



INDIANA UNIVERSITY

BORKENSTEIN
COURSE

Center for Studies
of Law in Action

Validation of the Standardized Field Sobriety Test Battery
At 0.08% Blood Alcohol Concentration

Jack Stuster

Validation of the Standardized Field Sobriety Test Battery at 0.08% Blood Alcohol Concentration

Jack Stuster, Anacapa Sciences, Inc., Santa Barbara, California

Objective: A field study was conducted to evaluate the accuracy of the Standardized Field Sobriety Test (SFST) battery to assist officers in making arrest decisions at blood alcohol concentrations (BACs) below 0.10%. **Background:** The SFST Battery was validated at 0.10% BAC in 1981, but since then many states have reduced statutory limits for driving while intoxicated to 0.08% BAC. **Methods:** During routine patrols, participating officers followed study procedures in administering SFSTs, scoring results, making arrest/no arrest decisions, and completing a data collection form for each of the 297 motorists evaluated during the study period. The officers' final step in each case was the administration of an evidentiary breath alcohol test. **Results:** Overall, officers' decisions were correct in more than 91% of the cases at the 0.08% BAC level. Cohen's kappa tests found all officers' scores to be within the categories of "substantial" and "near perfect" agreement, indicating low variance among the officers and a high degree of interrater reliability. **Conclusion:** The results of this study provide evidence of the validity of the SFST Battery as an accurate and reliable decision aid for discriminating between BACs above and below 0.08%. **Application:** The SFST Battery presently is used by law enforcement officers throughout the United States to help make roadside arrest decisions for impaired driving.

INTRODUCTION

The Standardized Field Sobriety Test (SFST) battery was developed to assist law enforcement officers in making roadside arrest decisions for alcohol-impaired driving at the 0.10% blood alcohol concentration (BAC) limit (Burns & Moskowitz, 1977; Tharp, Burns, & Moskowitz, 1981). The gradual adoption of 0.08% BAC as the statutory criterion for driving while impaired (DWI) offenses prompted an assessment of the SFST's utility to assist officers when making arrest decisions at BACs below 0.10%. The results of that assessment are presented here.

The SFST Battery

The three components of the SFST battery are described in the following paragraphs.

Horizontal gaze nystagmus test. Horizontal gaze nystagmus (HGN) is an involuntary jerking of the eye that occurs naturally as the eyes gaze to the side. Under normal circumstances, nystagmus

occurs when the eyes are rotated at high peripheral angles. However, nystagmus is exaggerated and may occur at lesser angles when a person is impaired by alcohol. An alcohol-impaired person also will have difficulty tracking a moving object smoothly (Aschan, 1958; Lehti, 1976; Pentilla, Tenhu, & Kataja, 1974; Wilkinson, Kime, & Purnell, 1974). In the HGN test, the officer observes the eyes of a suspect as the suspect follows a slowly moving object, such as a pen or small flashlight, horizontally with his or her eyes. The officer looks for three indicators of impairment in each eye: The eye cannot follow a moving object smoothly, jerking is distinct when the eye is at maximum deviation, and the angle of onset of jerking is within 45° of center. HGN also may indicate a rare and usually debilitating neurological condition or consumption of seizure medications, phencyclidine, certain inhalants, barbiturates, and other central nervous system depressants.

Walk and turn. The walk-and-turn and one-leg stand components of the SFST battery require

performance of physical tasks but are essentially "divided attention" tests. The tests require the person to listen to and follow instructions while performing simple physical movements. People who are impaired by alcohol have difficulty with tasks that require their attention to be divided between simple mental and physical exercises. In the walk-and-turn test, the motorist is directed to take nine steps, heel to toe, along a straight line. After taking the steps, the motorist must turn on one foot and return in the same manner in the opposite direction. The officer looks for eight indicators of impairment: The person cannot keep balance while listening to the instructions, begins before the instructions are finished, stops while walking to regain balance, does not touch the heel to the toe, steps off the line, uses arms to balance, makes an improper turn, or takes an incorrect number of steps.

One-leg stand. This test requires the motorist to stand with one foot approximately 6 inches (about 15 cm) off the ground while counting aloud by thousands (one thousand one, one thousand two, etc.) until instructed to lower the foot after 30 s. The officer looks for four indicators of impairment: swaying while balancing, using arms to balance, hopping to maintain balance, and putting the foot down before instructed to do so.

METHOD

Field experience with the SFST battery at 0.10% BAC led to the hypothesis that modified scoring of test results could provide officers with the information necessary to make accurate roadside arrest decisions at lower BACs. The hypothesized changes to the procedures were simple: The exhibition of two or more clues on the one-leg stand or walk-and-turn tests would indicate a BAC of at least 0.08%, rather than 0.10%, which had been established by the original studies. Similarly, the exhibition of four HGN clues would indicate a BAC of 0.08% or greater, rather than 0.10%. (Note: The eyes are scored separately with a total of six clues possible.) The only other deviation from the existing procedures would be to use the exhibition of two HGN clues as an indication of BACs greater than zero but below 0.08%.

Site Selection

Several factors constrained the site selection

process and limited the possible candidates for participation in this study. First, at the time the project was conducted, California, Oregon, and Utah were the only states that met both of the BAC-related site selection criteria – namely, a 0.08% BAC limit for DWI and a zero-tolerance law for drivers under 21 years of age. Second, it was important to restrict the data collection period, to the extent possible, because it was believed that an extremely long period might result in officers deviating from the study procedures. Strict adherence to study procedures was considered essential to ensuring the internal validity of the study.

All participating officers in the 1981 study were trained by the researchers to administer and score the experimental test battery. However, more than 300,000 officers have received SFST training since 1981. For this reason, it was determined that only officers who had previously received SFST training from a certified instructor could participate in the current study, whereas in 1981 novice officers were trained. The purposes of the stipulation were to ensure that experimental conditions accurately reflected the operational condition of widespread use of the SFSTs by law enforcement officers and to avoid confounding study results with the effects of differential officer skill and experience levels in SFST administration and scoring. The city of San Diego, California, was identified as the community that best satisfied the site selection criteria.

Dependent Measures

The experiment plan focused on obtaining data about adult motorists who were suspected of exceeding the statutory limit of 0.08% BAC and drivers under 21 who were suspected of exceeding the zero-tolerance legal limit of 0.00%. The utility of the SFSTs to discriminate at 0.08% and 0.04% BAC could not be tested without data from drivers who had BACs over and under these values. BACs and officers' arrest decisions are the only appropriate criteria for evaluating the accuracy of the SFSTs. Measures of impairment are irrelevant because performance of the SFSTs must be correlated with BAC level, not with driving performance. BAC provides an objective and reliable measure that states have recognized as presumptive and/or per se evidence of impairment, depending on the statute. To obtain these criterion measures, it was determined that all drivers who were administered the SFST battery

must be tested for BAC, regardless of the results of the SFSTs or the officers' arrest decisions.

Procedures

The requirement for an agency to modify its established methods of operation to accommodate research requirements usually is somewhat negotiable in a traffic safety study. However, deviations from study procedures were not allowed in this case. All participating officers were required to abide by the established procedures. In particular, no tests other than the SFSTs could be administered to drivers, and officers were required to perform the tasks in the prescribed sequence.

The procedures were listed as a series of six numbered steps on the data collection forms used in the field study. Officers were instructed to perform their normal patrol duties and to follow the procedures whenever an adult driver was suspected of being alcohol impaired or a driver under 21 was suspected of having a BAC greater than zero. In practice, officers administered the SFSTs to all motorists who exhibited any objective behavior or other driving cue associated with the effects of alcohol, even if impairment was not evident. A breath, blood, or urine test was administered to all motorists who performed the SFSTs, but only after the officer had scored and recorded the SFST test results, BAC estimate, and arrest/no arrest decision on the form. The data collection form structured the procedure by presenting all officer actions as a series of numbered steps. Requiring officers to record the times of BAC estimates and BAC tests on the form provided verification that officers' estimates were not influenced by the results of chemical tests performed later. The 7 participating officers received 4 hr of training in the administration of study procedures, and their performance was verified by a certified SFST instructor. Also, the officers and police managers signed memoranda of agreement affirming that they would abide by the established procedures during the study period. In addition, project staff periodically accompanied officers on patrol to observe the data collection effort.

Legal Issues

In some states, including California, officers have the right to administer a breath test to a driver who has exhibited any objective sign of alcohol consumption. Compliance is mandatory if the

officer can articulate a reasonable suspicion of the motorist having consumed alcohol (such as the odor of an alcoholic beverage). SFSTs were administered only to drivers who exhibited some objective DWI cue; thus there were no legal or methodological obstacles to obtaining BAC data, even from motorists whose SFST performance was acceptable. A field breath test was conducted as the final step after the SFST procedure was completed, which is the de facto procedure followed by most officers who are equipped with field breath testing devices.

Materials

A pen, pencil, or small flashlight usually is used by officers as a stimulus or target for the HGN test, but a finger is equally effective; that is, no special equipment was required to administer the three components of the SFST battery. However, officers were equipped with evidentiary-quality, handheld breath testing devices with digital displays to obtain breath test data.

The data collection form was designed to be as simple to complete as possible in order to minimize the intrusion of study requirements on the workload and safety of participating officers. The form also was designed to guide the officers in the administration of the SFSTs, to facilitate standardized and systematic scoring of the tests, and to both encourage and provide assurances that officers had followed the study procedures. Most important, it was essential that officers would conduct a breath test and record the driver's BAC as the final steps of the process; that is, actual BACs were to be entered on the form, and the time of the breath test recorded, only after SFST results, the officer's BAC estimate, and the arrest/no arrest decision had been recorded.

RESULTS

The 7 participating officers from the San Diego Police Department's alcohol enforcement unit completed a total of 298 data collection forms during the study period; 1 case was eliminated from analysis because the driver refused to submit to any form of BAC testing. Officer compliance with study procedures and motivation to participate in the study remained high throughout the data collection period. The officers' practice of administering the SFSTs to all motorists who exhibited

any behavior associated with the effects of alcohol, even if impairment was not evident, provided additional low-BAC drivers to the sample.

Evaluation of SFST Accuracy

Decision matrices were constructed to describe the four possible combinations of the two variables of interest: arrest decisions based on estimated BACs and actual BACs above and below the criterion level. Table 1 presents the first matrix, with the four major cells representing the four possible decisions at 0.08% BAC. The numbers in the major cells are the frequencies of each type of decision out of the 297 SFST administrations. The two italicized cells in the matrix represent officers' correct decisions: (a) 212 motorists who officers estimated to have BACs $\geq 0.08\%$ and later were found to have BACs $\geq 0.08\%$ by BAC testing (by breath, blood, or urine analysis) and (b) 59 motorists who officers estimated to have BACs $< 0.08\%$ and later tested below 0.08%. The table also shows the incorrect decisions: 22 motorists who officers estimated to have BACs $> 0.08\%$ but who later were found to have BACs below that level (false positives) and 4 motorists who officers estimated to have BACs below 0.08% and who later tested $\geq 0.08\%$ (false negatives). Overall, officers' decisions were correct in more than 91% of the cases.

Cohen's kappa test results and a categorical interpretation of agreement also are presented in Table 1. According to Landis and Koch (1977), a κ value of .21 to .40 indicates "fair agreement," .41 to .60 indicates "moderate agreement," .61 to .80 indicates "substantial agreement," and .81 to 1.0 indicates "almost perfect agreement." The κ value of .7628 for all SFST administrations reflects substantial overall agreement between all the officers' decisions (i.e., $\geq 0.08\%$ or $< 0.08\%$ BAC) and the actual BACs that were measured later. Table 2 presents a decision matrix, the associated

κ score, and the agreement category for each of the 7 participating officers. Officers' individual κ scores ranged from .611 to 1.0; all scores are within the categories of "substantial" and "near perfect" agreement, indicating low variance among the officers and a high degree of interrater reliability.

The officers' arrest decisions based on estimated BACs were inconsistent with the number of HGN clues observed in 7 of the false positive and 2 of the false negative cases; that is, accuracy would have been better had the officers not deviated from the scoring guidelines in these cases. Further, 7 of the false positives had measured BACs of 0.07% or greater, but less than 0.08%, all of which are within the margin of error of the testing devices. A 10th case was a juvenile with a measured BAC of .069%, which rendered the difference between estimated and measured BACs irrelevant in a zero-tolerance jurisdiction; that is, it was a correct arrest decision despite the overestimation of the juvenile's BAC. In addition, two of the motorists with measured BACs of 0.07% were arrested for DWI because the officers believed that they were too impaired to be permitted to drive. Finally, an additional motorist with an estimated BAC of 0.08% and a measured BAC of 0.05% was found to be a psychiatric patient, which helped to explain her erratic behavior, poor SFST performance, and obvious impairment.

Further analyses were performed to explore methods for combining the results of the three component tests. Only the 261 cases that included test results for all three component tests could be used in this analysis. Of those drivers, 73 were found to have BACs $< 0.08\%$ and 188 had measured BACs $\geq 0.08\%$. In 162 of the 188 cases (86%), all three component SFSTs were unanimous in their predictions.

Figure 1 presents a Venn diagram that illustrates the contributions of the three tests to the 14%

TABLE 1: Decision Matrix at 0.08% BAC for All Cases (7 Officers, $n = 297$)

Measured BAC	Estimated BAC		Kappa Score	Agreement $p < .05$
	$\geq 0.08\%$	$< 0.08\%$		
$\geq 0.08\%$	212	4	.7628	Substantial
$< 0.08\%$	22	59		

Note. Numbers in italics indicate officers' correct decisions.

TABLE 2: Decision Matrices at 0.08% BAC for Each Officer

Officer	<i>n</i>	Measured BAC	Estimated BAC		Kappa Score	Agreement <i>p</i> < .05
			≥0.08%	<0.08%		
1	58	≥0.08%	38	1	.6110	Substantial
		<0.08%	8	11		
2	31	≥0.08%	24	0	1.0000	Perfect
		<0.08%	0	7		
3	69	≥0.08%	39	2	.7859	Substantial
		<0.08%	5	23		
4	47	≥0.08%	41	0	.6356	Substantial
		<0.08%	3	3		
5	44	≥0.08%	37	0	.8078	Almost perfect
		<0.08%	2	5		
6	25	≥0.08%	16	0	.7191	Substantial
		<0.08%	3	6		
7	23	≥0.08%	17	1	.7444	Substantial
		<0.08%	1	4		

Note. Numbers in italics indicate officers' correct decisions.

of cases in which a discrepancy occurred. The figure shows there were 162 cases with BACs ≥ 0.08 in which all three SFSTs indicated a BAC of ≥ 0.08 and 26 cases in which one or more tests disagreed (the numbers inside the circles). A single test indicated a BAC $< 0.08\%$ in 17 of the cases ($8 + 2 + 7$), and two tests were involved in nine of the cases ($1 + 1 + 7$). There were no cases in which all three tests predicted incorrectly.

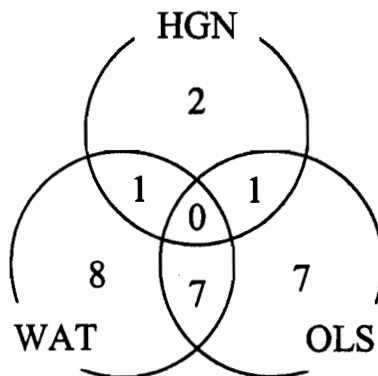


Figure 1. Venn diagram of 188 cases $\geq 0.08\%$ BAC, showing the 26 cases in which all three tests do not agree. The 162 cases in which all three tests agree are represented by the area outside the three circles. (HGN = horizontal gaze nystagmus test, WAT = walk and turn test, OLS = one-leg stand test.)

The horizontal gaze nystagmus (HGN) test was about four times less likely to be the source of a discrepancy than the other two tests. Only 2 of the single-test discrepancies were attributable to HGN results, compared with 8 cases for the walk and turn test and seven cases for the one-leg stand test. Overall, the HGN test was involved in only 4 of the discrepancies, compared with 16 cases for the walk and turn and 15 cases for the one-leg stand.

The question of the SFST battery's accuracy in discriminating BACs above and below 0.04% is addressed by the decision matrix presented in Table 3; the italicized cells of the matrix again represent correct decisions based on SFST results. The table shows that officers estimated motorists' BACs to be equal to or greater than 0.04% but under 0.08% in 54 cases, and in 51 of those cases their estimates were found to be correct by subsequent breath, blood, or urine testing. The table also shows that officers estimated that 29 motorists had BACs below 0.04%, and in 15 of those cases their estimates were found to be correct by subsequent testing. Overall, officers were accurate in 80% of the cases when discriminating between motorists who were above 0.04% but below 0.08% BAC. However, kappa test results indicated only "moderate" agreement between officers' decisions and measured BACs ($\kappa = .5061$).

TABLE 3: Decision Matrix at 0.04% BAC (All Cases <0.08% BAC; 7 Officers, *n* = 83)

Measured BAC	Estimated BAC		Kappa Score	Agreement <i>p</i> < .05
	≥0.04%, <0.08%	<0.04		
≥0.04%, <0.08%	51	14	.5061	Moderate
<0.04%	3	15		

Note. Numbers in italics indicate officers' correct decisions.

DISCUSSION

The research described here found that the SFST battery accurately and reliably assists officers in making DWI arrest decisions at 0.08% BAC. The primary implication of the study results is that the SFST battery is a valid decision aid for making roadside DWI assessments.

The horizontal gaze nystagmus test was found to be the most predictive of the three component tests, but correlations with measured BACs were higher when the results of all three tests were combined, as reported earlier. The implications of this study result are that all components of the SFST battery should be administered when possible or practical. However, the data indicate that the HGN test alone can provide valid indications to support officers' arrest decisions at 0.08% BAC.

Note About the Acceptability of the HGN Test

The horizontal gaze nystagmus (HGN) test is considered by many law enforcement officers to be a foolproof technique (sometimes called a "silver bullet") that provides indisputable evidence of alcohol in a motorist's system. The normal variation in human physical and cognitive capabilities, and the effects of alcohol tolerance, result in uncertainties when arrest decisions are made exclusively on the basis of performance tests. These uncertainties have resulted in large proportions of DWI suspects being released rather than detained and transported to another location for evidentiary chemical testing. This is important because experienced drinkers often can perform physical and cognitive tests acceptably with a BAC greater than 0.10%. However, even the most experienced drinkers cannot conceal the physiological effects of alcohol from an officer skilled in HGN test administration, because nystagmus is an involuntary

reaction over which an individual has absolutely no control. Many individuals, including some judges, believe that the purpose of a field sobriety test is to measure driving impairment. For this reason, they tend to expect tests to possess "face validity" – that is, the appearance of being related to actual driving tasks. Tests of physical and cognitive abilities, such as balance, reaction time, and information processing, have face validity, to varying degrees, based on the involvement of these abilities in driving tasks; that is, the tests seem to be relevant "on the face of it." HGN lacks face validity because it does not appear to be linked to the requirements of driving a motor vehicle. The reasoning is correct, but it is based on the incorrect assumption that field sobriety tests are designed to measure driving impairment.

Driving a motor vehicle is a complex activity that involves many tasks and operator capabilities that would be difficult, if not impossible, to measure at roadside. The constraints imposed by field conditions were recognized by the developers of the SFST battery. As a consequence, they pursued the development of tests that would provide valid indications of a driver's BAC, rather than indications of driving impairment. The link between BAC and driving impairment is a separate issue, involving entirely different research methods. Those methods have found driving to be impaired at BACs as low as 0.02%, with a sharp increase in impairment at about 0.07% (Moskowitz & Robinson, 1988; Stuster, 1997). Thus, SFST results help officers to make accurate DWI arrest decisions even though SFSTs do not directly measure driving impairment. HGN's apparent lack of face validity to driving tasks is irrelevant because the objective of the test is to discriminate between drivers above and below the statutory BAC limit, not to measure driving impairment. Throughout the United States, DWI laws permit arrest decisions